A MODIFIED LEMPEL–ZIV WELCH SOURCE CODING ALGORITHM FOR EFFICIENT DATA COMPRESSION

1 MAHMOUD A. SMADI, 2 QASEM ABU AL-HAIJA
1, 2 Electrical Engineering Department, King Faisal University, Saudi Arabia, Alhasa, 31982, P. O. Box 380
E-mail: 1 msmadi@kfuf.edu.sa, 2 qalhaija@kfuf.edu.sa

ABSTRACT

Lempel–Ziv Welch (LZW) algorithm is a well-known powerful data compression algorithm created by Abraham Lempel, Jacob Ziv, and Terry Welch. The algorithm is designed to be fast to implement but is not usually optimal because it performs only limited analysis of the data. A modified LZW algorithm on source coding will be proposed in this paper to improve the compression efficiency of the existing algorithms. Such method is to be implemented with appropriate modifications that gives the best performance and satisfies the requirements of the applications.

Keywords: LZW algorithm, substitution compression, encoding/decoding process.

1. INTRODUCTION

Data compression seeks to reduce number of bits used to store or transmit information. It encompasses a wide variety of software and hardware compression techniques [1], which can be so unlike one another except that they compress data. These techniques for data compression can ease a variety of problems in storing and transmitting large amount of data [2].

Because of the tremendous amount of digital data used today in several applications in digital data systems, i.e. video-on-demand, word processing programs, digital signal processing etc., the channel bandwidth and the disk drive that looked gigantic become inadequate for most applications. To provide transmission or storage facilities for the data, we need an additional communication lines for transmission or disk drives for storage. In addition to these solutions, the auxiliary devices such as modems, multiplexers ... etc. have been continuously upgraded to permit higher data transfer capability [3].

The above ordinary solutions used in transmitting or storing large amount of data require an additional increase in organization equipment and operating costs. One method that can be employed to improve a portion of data storage and information transfer problems is to seek about sophisticated algorithms to search data for redundancy [2]. This redundancy can be removed from the original data by compression algorithms; consequently, the resulting data is smaller than the original data (reduction the quantity of data or information). After that, the compressed data enters to the transmission medium and then it is expanded to its original format at a sink location or we store it on disk drives and then it is expanded when it is needed.

Different compression performance will be obtained by applying the context in different way. For example, a compression method using fixed dictionary has a high speed, but the compression ratio is worse than that of methods using dynamic dictionary. Hence, we search in this paper to propose a variant matching procedure to improve compression performance based on LZW algorithm.

2. DATA COMPRESSION TECHNIQUES

There are many classifications for data compression schemes. These classifications depend on the way in which the techniques treat the text to be compressed. A text is constructed as a group of characters, which are arranged in a random sequence. The probability of occurrence of a character in a text is not the same for all characters. For example, in a typical English language the probability of occurrence of the space and vowel characters (e, o, a, i, u) is much higher than that of other characters such as z or the question mark character. By utilizing this fact, we can assign short codes for most frequently occurring set of characters, while long codes are assigned for seldom occurring set of characters. This method of data compression depends on the frequency of